MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DJILLALI LIABES SIDI BEL ABBES

FACULTÉ DES SCIENCES DE L'INGÉNIEUR

RESUMÉ DE THÈSE DE DOCTORAT

Nom & Prénom(s)	TOUMOUH Adil
E-mail (obligatoire)	toumouh@gmail.com
Spécialité	Informatique
Titre	Ingénierie des Ontologies : Les Corpus Parallèles et le Word Space Model dans l'Enrichissement des Ontologies Multilingues.
Date de soutenance	27 novembre 2013
Nom, prénom(s) et grade de l'encadreur	LEHIRECHE Ahmed, MCA

Résumé :

Les ressources sémantiques multilingues sont devenues une nécessité face aux nouvelles exigences imposées par la diversité linguistique et culturelle des consommateurs de l'information et qui sont issus de diverses nations.

Cette thèse entre dans le cadre de l'enrichissement des ontologies multilingues. L'approche développée utilise les corpus parallèles comme la seule et l'unique donnée en entrée. Cette masse de données est représentée à l'aide du modèle Espace-de-mots. Pour remédier au problème de tractabilité des matrices à haute dimension, nous utilisons le *Random Indexing* à fin de réduire la dimension du modèle algébrique.

L'enrichissement que nous proposons agit au niveau terminologique. Trouver des synonymes dans d'autres langues est accomplie en opérant par des traductions au niveau des termes. Le taux de réussite de l'ordre de 90% est plus que satisfaisant.

Dans une seconde étape, nous proposons le modèle interlingua pour les langues les moins populaires et qui sont pauvres en matière de corpus parallèles. Comparé au modèle Espace-de-mots direct, le modèle Espace-de-mots interlingua (indirect) enregistre un taux de réussite de 85%. Alors, nous explorons deux caractéristiques : la taille des corpus et la dimension du modèle Espace-de-mots, afin d'identifier les conditions dans lesquelles le modèle interlingua peut rivaliser avec le modèle direct.

Mots clés : Ontologie multilingue, Enrichissement, Synonymie, Corpus Parallèle, Espace-de-Mots, Random Indexing.

ملخص

أَصْبَحتْ مصادر دلالات الألفاظ، المتعدِّدة اللّغات، ضرورة في ضوء المتطلبات الجديدة التي يفرضها التنوّع اللغوي والثقافي لمستهلكي المعلوماتِ من مختلف الأُمم.

تندرج هذه الأطروحة في مجال إثراء "ontology" المتعدّدة اللغات. التقنيّة المطوّرة في هذه الرسالة تعتمد على إستخدام المجاميع الموازية، باعتبارها المعطيات الوحيدة في النظام. يُمثِّلُ هذا الكمّ الهائل من البيانات باستخدام نموذج "فضاء الكلمات". لمعالجة مشكلة المصفوفات ذات الأبعاد العالية، نستخدم الفهرسة العشوائية لتقليل حجم النموذج الجبري.

الإثراء المقترح يعمل على المستوى الإصطلاحي فقط، ولإيجاد مرادفات في اللّغات الأخرى نقوم بترجمة الألفاظ؛ تُعتبر نسبة النّجاح المقدّرة بحوالي 90٪ أكثر من مرضية.

كإسهامٍ ثانٍ، اقترحنا نموذج اللّغة الوسيطة " interlingua model " للّغات الأقلّ شعبية والفقيرة من ناحية المجاميع الموازية. قمنا بمقارنةٍ نموذج "فضاء الكلمات" المباشر مع نموذج "فضاء الكلمات" باستعمال اللّغة الوسيطة، فتبيَّن لنا أنّ نسبة نجاح هذا الأخير لا تتجاوز 85٪. كتجربة جديدة، استخدمنا الخاصّيتين : حجم المجاميع وأبعاد "فضاء الكلمات"، وذلك لتحديد الشّروط التي تسمح للنّموذج المبني على اللّغة الوسيطة بمنافسة النموذج المباشر من حيث النتائج.

كلمات مفتاحيه : إثراء، " ontology " متعدّدة اللغات، الترادف، المجاميع الموازية، "فضاء الكلمات"، الفهرسة العشوائية.

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DJILLALI LIABES SIDI BEL ABBES

FACULTÉ DES SCIENCES DE L'INGÉNIEUR

RESUMÉ DE THÈSE DE DOCTORAT